

PENERAPAN DATA MINING DENGAN METODE KALSIFIKASI MENGUNAKAN *DECISION TREE* DAN REGRESI

Popy Meilina^{1,*}

¹Jurusan Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Jakarta

Jl. Cempaka Putih Tengah 27 Jakarta 10510

*E-mail: popy.meilina@gmail.com

Diterima: 31 Agustus 2014

Direvisi: 23 September 2014

Disetujui: 11 Nopember 2014

ABSTRAK

Konversi Bahan Bakar Minyak (BBM) ke Bahan Bakar Gas (BBG) yang dilakukan pemerintah bermula atas melimpahnya gas di bumi Indonesia. kabupaten Cilacap melakukan survey terhadap masyarakat untuk menentukan masyarakat mana yang lebih diprioritaskan untuk mendapatkan bantuan pemberian kompor gas beserta tabungnya. Berdasarkan survey pada tahun 2010 di Kabupaten Cilacap Kecamatan Kroya terdapat 17 kelurahan dengan 28.131 data dengan 31 field atau kolom. Data diolah dengan beberapa tahapan yaitu : pengecekan data, integrasi data, target data, preprocessing, visualisasi, proses data mining, dan pengetahuan. Pada proses data mining dilakukan proses decision tree, regresi, dan k means untuk clustering yang ditampilkan dalam bentuk scatter plot. Berdasarkan decision tree maka diperoleh data kelurahan yang menjadi prioritas untuk diberi bantuan yaitu : Kelurahan Bajing Kulon, Kedawung, Pekuncen, dan Pesanggrahan. Atribut yang digunakan untuk proses regresi yaitu per kelurahan dengan variabel x yaitu pendidikan atau pekerjaan dan variabel y yaitu penghasilan.

Kata kunci: data mining, decision tree, regresi

ABSTRACT

A conversion from oil fuel to gas fuel that done by Indonesian Government are based on the overwhelming gas in Indonesia. Cilacap District has done a survey to people that live there in order to determine which people that would be consider to be prioritized to get subsidiary gas stove and gas tube. Based on survey in 2010 at Kroya an administrative area on Cilacap District, that have 17 villages with 28.131 data and 13 field apply. Data was processed with several stages, which is: data checking, data integration, data target, preprocessing, visualize and data mining process. In data mining process, there were comparison of three methods that are Decision Tree, Regression, and K-Means Clustering which visualize in a form of Scatter Plot. Results with decision tree method gain that subsidiary priorities are following villages; Bajing Kulon, Kedawung, Pekuncen and Pesanggrahan. While on Regression method using education and job status for x variable and income for y variable.

Keywords : data mining, decision tree, regression

PENDAHULUAN

Indonesia dengan kekayaan gas alam yang sangat besar, seharusnya pengadaan energi gas dapat dimaksimalkan untuk memenuhi kebutuhan dalam negeri. Namun masyarakat masih sedikit yang memanfaatkan

kekayaan ini. Masyarakat masih memandang Bahan Bakar Minyak (BBM) sebagai energi yang utama dalam kegiatan sehari-harinya.

Jumlah subsidi untuk BBM yang sangat besar, dan kondisi Indonesia yang sudah menjadi net importir minyak menjadi beberapa alasan bagi Indonesia untuk segera melakukan

program konversi BBM ke Bahan Bakar Gas (BBG).

Diawali dengan keinginan pemerintah untuk mengkonversi Bahan Bakar Minyak (BBM) ke Bahan Bakar Gas (BBG), sehingga BUMN (Badan Usaha Milik Negara) khususnya di kabupaten Cilacap pada tahun 2009 berusaha untuk memaksimalkan keinginan pemerintah yang kemudian akan memberikan secara cuma-cuma kompor gas dan tabung gas untuk golongan yang tidak mampu, dan kemudian BUMN pun melakukan *survey* secara menyeluruh untuk mendapatkan data yang *valid* kelompok masyarakat mana yang diprioritaskan pertama untuk mendapatkan bantuan tersebut.

Berdasarkan data *survey* yang dilakukan BUMN Peneliti bermaksud menggunakannya untuk mendapatkan pengetahuan sehingga bisa dimanfaatkan oleh berbagai pihak untuk kepentingan tertentu khususnya dalam hal ini pemerintah terkait. Analisa yang dilakukan menggunakan Data Mining berdasarkan dari karakteristik data yang diperoleh seperti Pekerjaan, Pendidikan, Penghasilan dan atau Pengeluaran sehingga hasil yang didapat diharapkan lebih tepat sasaran.

Data Mining (Penambangan data) merupakan bidang ilmu multidisplin, menggambarkan area-area kerja yang termasuk didalamnya adalah teknologi basis data, pembelajaran mesin, statistik, pengenalan pola, pengambilan informasi, jaringan saraf tiruan, sistem berbasis pengetahuan, kecerdasan buatan, komputasi kinerja-tinggi, dan visualisasi data [Han & Kamber, 2006].

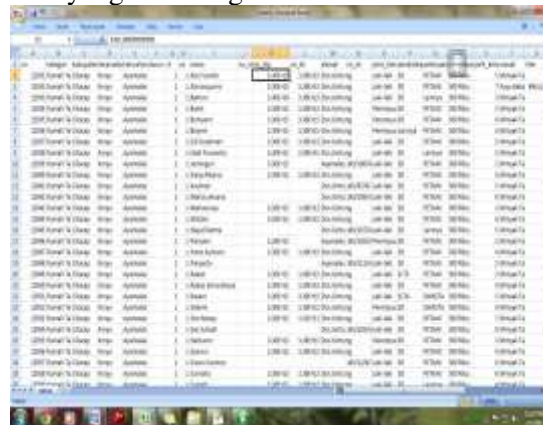
Ada banyak metodologi Data Mining, salah satu yang populer adalah Pohon Keputusan (*Decision Tree*). Pohon keputusan merupakan salah satu metode klasifikasi yang sangat menarik yang melibatkan konstruksi pohon keputusan yang terdiri dari node keputusan yang di hubungkan dengan cabang-cabang dari simpul akar sampai ke node daun (akhir). Pada node keputusan atribut akan diuji, dan setiap hasil akan menghasilkan cabang. Setiap cabang akan diarahkan ke node lain atau ke node akhir untuk menghasilkan suatu keputusan (Larose, 2005).

Menurut Maimon (2005), pohon keputusan adalah salah satu metode klasifikasi yang dinyatakan sebagai partisi rekursif. Pohon keputusan terdiri dari node yang

membentuk pohon yang berakar, semua node memiliki satu masukan. Node yang keluar disebut node tes. Node yang lain disebut node keputusan atau sering disebut node daun. Setiap simpul internal membagi dua atau lebih sub-ruang sesuai dengan kategori atribut dan akan dipartisi sesuai dengan nilai kategori kasus. Kasus-kasus tersebut membentuk pohon keputusan yang menghasilkan problem solving.

Klasifikasi data mining adalah penempatan objek-objek ke salah satu dari beberapa kategori yang telah ditetapkan sebelumnya. Klasifikasi banyak digunakan untuk memprediksi kelas pada suatu label tertentu, yaitu dengan mengklasifikasi data (membangun model) berdasarkan training set dan nilai-nilai (label kelas) dalam mengklasifikasikan atribut tertentu dan menggunakannya dalam mengklasifikasikan data yang baru. Pohon keputusan biasanya digunakan untuk mendapatkan informasi untuk tujuan pengambilan sebuah keputusan. Pohon keputusan dimulai dengan sebuah root node (titik awal) yang digunakan oleh user untuk mengambil tindakan. Berdasarkan node root ini, user memecahkan leaf node sesuai dengan algoritma decision tree. Regresi linier adalah metode statistika yang digunakan untuk membentuk model hubungan antara variabel terikat (*dependen*) dengan satu atau lebih variabel bebas (*independen*). Apabila banyaknya variabel bebas hanya ada satu, disebut sebagai regresi linier sederhana, sedangkan apabila terdapat lebih dari 1 variabel bebas, disebut sebagai regresi linier berganda [Xin Yan dan Xiao Gang Su, 2009].

Berikut di bawah ini adalah tampilan data yang Peneliti gunakan.



No	Nama	Jenis Kelamin	Usia	Pendidikan	Pekerjaan	Penghasilan	Pengeluaran	Status
1	10001	Pria	35	S1	Manajemen	1000000	500000	Menikah
2	10002	Wanita	25	S1	Manajemen	1000000	500000	Menikah
3	10003	Pria	30	S1	Manajemen	1000000	500000	Menikah
4	10004	Wanita	20	S1	Manajemen	1000000	500000	Menikah
5	10005	Pria	30	S1	Manajemen	1000000	500000	Menikah
6	10006	Wanita	25	S1	Manajemen	1000000	500000	Menikah
7	10007	Pria	30	S1	Manajemen	1000000	500000	Menikah
8	10008	Wanita	25	S1	Manajemen	1000000	500000	Menikah
9	10009	Pria	30	S1	Manajemen	1000000	500000	Menikah
10	10010	Wanita	25	S1	Manajemen	1000000	500000	Menikah
11	10011	Pria	30	S1	Manajemen	1000000	500000	Menikah
12	10012	Wanita	25	S1	Manajemen	1000000	500000	Menikah
13	10013	Pria	30	S1	Manajemen	1000000	500000	Menikah
14	10014	Wanita	25	S1	Manajemen	1000000	500000	Menikah
15	10015	Pria	30	S1	Manajemen	1000000	500000	Menikah
16	10016	Wanita	25	S1	Manajemen	1000000	500000	Menikah
17	10017	Pria	30	S1	Manajemen	1000000	500000	Menikah
18	10018	Wanita	25	S1	Manajemen	1000000	500000	Menikah
19	10019	Pria	30	S1	Manajemen	1000000	500000	Menikah
20	10020	Wanita	25	S1	Manajemen	1000000	500000	Menikah
21	10021	Pria	30	S1	Manajemen	1000000	500000	Menikah
22	10022	Wanita	25	S1	Manajemen	1000000	500000	Menikah
23	10023	Pria	30	S1	Manajemen	1000000	500000	Menikah
24	10024	Wanita	25	S1	Manajemen	1000000	500000	Menikah
25	10025	Pria	30	S1	Manajemen	1000000	500000	Menikah

Gambar 1. Tampilan sumber data (data mentah) hasil survey Pihak BUMN pada Kabupaten Cilacap Kecamatan Kroya

Seperti terlihat pada gambar 1 terdapat beberapa field yang dijadikan pedoman bagi pihak BUMN yaitu: Kategori, kabupaten, kecamatan, kelurahan, dusun, rt, rw, nama, no id, no ktp, no kk, alamat, no sk, jenis kelamin,

pendidikan, pekerjaan, pengeluaran, jumlah keluarga, masak, liter, dan lain-lain. Data yang ada yaitu 28.131 survey data dengan 31 field atau kolom.

Dibawah ini adalah field-field yang diambil dari data kependudukan Kecamatan Kroya yang ada pada data MS. Excell.

Tabel 1 Tabel Data Penduduk

Atribut	Keterangan
Kategori	Didalamnya mencakup data yang terdiri dari : Rumah Tangga tetap, Usaha Kecil Menengah (UKM), Rumah Tangga Musiman
Kelurahan	Mencakup 17 macam Kelurahan yang ada pada Kecamatan Kroya
Jenis Kelamin	Merupakan jenis kelamin penduduk yang bersangkutan
Pendidikan	Merupakan jenjang pendidikan akhir yang diselesaikan oleh penduduk yang bersangkutan
Pekerjaan	Jenis pekerjaan yang dilakoni oleh setiap penduduknya
Pengeluaran	Seberapa banyak pengeluaran yang dikeluarkan oleh setiap keluarga perbulannya
Jumlah keluarga	Banyaknya jumlah anggota setiap keluarga
Masak	Dengan menggunakan apa setiap penduduknya memasak
Liter	Jumlah liter yang dihabiskan oleh penduduknya jika memasak dengan menggunakan kompor minyak
Rupiah	Jumlah pengeluaran untuk minyak tanah yang harus dibeli oleh penduduk yang memasak dengan menggunakan kompor minyak.

METODE

Sumber data untuk penelitian merupakan data hasil survey yang dilakukan oleh BUMN Kabupaten Cilacap Kecamatan Kroya pada tahun 2009. Data yang digunakan dalam penelitian ini sebanyak 28.131 data pada Kecamatan Kroya yang terdiri dari 17 kelurahan. Langkah-langkah yang digunakan dalam penelitian ini adalah sebagai berikut :

1. Pengecekan Data (Proses Pembersihan Data)

Dalam Penelitian kali ini dicari hubungan beberapa atribut dari data yang ada. Karena tidak semua tabel yang digunakan maka perlu dilakukan pembersihan data yang akan diolah benar-benar relevan dengan yang dibutuhkan. mengapa data perlu di bersihkan sebelum diproses ? Hal ini terjadi karena biasanya data yang akan digunakan belum baik, penyebabnya antara lain :

- Incomplete : kekurangan nilai-nilai pada atribut atau atribut tertentu lainnya.

- Noisy : berisi kesalahan atau nilai-nilai outlier yang menyimpang yang tidak sesuai dengan data yang lainnya
- Inconsisten : ketidakcocokan dalam penggunaan kode atau nama. Disini kualitas data yang baik didasarkan oleh keputusan yang baik dan data warehouse memerlukan integrasi kualitas data yang konsisten.

Pada tahap ini proses Pembersihan data menggunakan dua tools yaitu, Microsoft Excell dan Orange.

2. Proses Pengintegrasian Data

Integrasi data merupakan penggabungan data dari berbagai *database* ke dalam satu *database* baru. Disini proses mengubah data atau konversi data dalam bentuk numerik. Sama seperti tahap sebelumnya, pada tahap ini proses pengintegrasian data menggunakan dua tools yang saling mendukung satu sama lain, yaitu Microsoft Excell dan Orange.

3. Target Data

Tahap ini merupakan penjelasan mengenai data yang siap digunakan dalam proses data mining, dimana target data menjadi input untuk proses-proses selanjutnya.

4. Preprocess

Sebelum diproses data mining sering kali diperlukan *preprocessing*. *Data preprocessing* menerangkan tipe-tipe proses yang melaksanakan data mentah untuk mempersiapkan proses prosedur yang lainnya. Tujuannya *preprocessing* dalam data mining adalah mentransformasi data ke suatu format yang prosesnya lebih mudah dan efektif untuk kebutuhan pemakai, dengan indikator sebagai berikut :

- Mendapatkan hasil yang lebih akurat
- Pengurangan waktu komputasi untuk *large scale problem*
- Membuat nilai data menjadi lebih kecil tanpa merubah informasi yang dikandungnya.
- Terdapat beberapa metode yang digunakan untuk *preprocessing* seperti :
 - a. *Sampling*, menyeleksi subset representatif dari populasi data yang besar.
 - b. *Diskretisasi*, Bagian dari data reduksi tetapi memiliki arti penting tersendiri, terutama untuk data numerik.
 - c. *Remove missing*
 - d. *Continueize*
 - e. *Impute*
 - f. *Feature selection*

5. Visualisasi

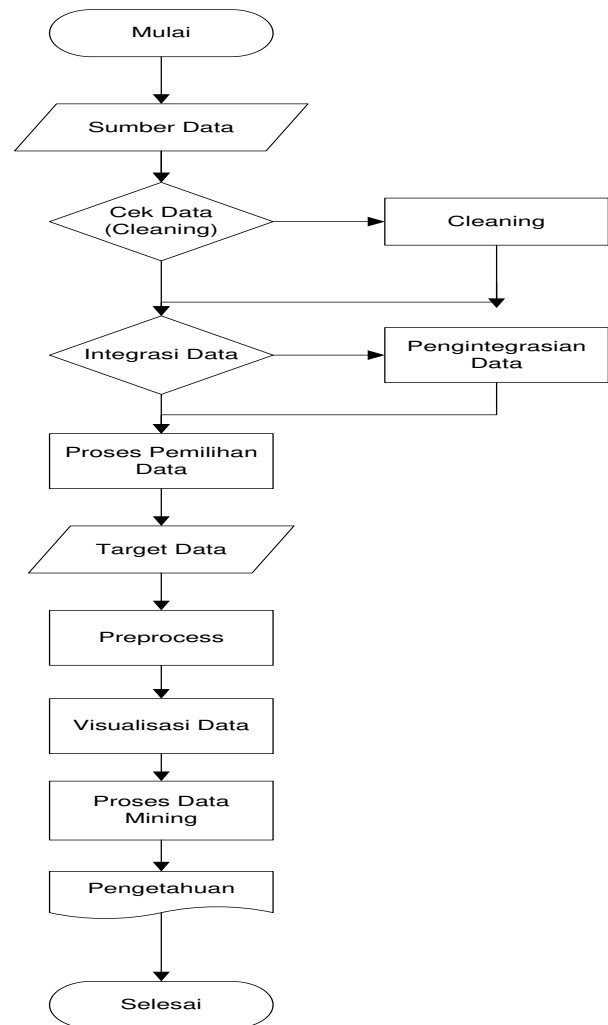
Tahap selanjutnya adalah visualisasi data, yaitu melihat data dengan menggunakan penjabaran secara grafis seperti Histogram, distribution diagram, diagram titik atau penjabaran secara lokasi seperti mean, median, modus, kuartil dan persentil.

6. Proses Data Mining

Tahap ini merupakan tahap pembentukan *decision tree*, regresi, dan klasifikasi K-means. Proses yang dilakukan pada *decision tree* menggunakan *software orange* dan proses untuk mencari regresi serta K-Means menggunakan *software matlab*.

7. Pengetahuan

Diharapkan setelah melalui setiap tahapan maka akan dihasilkan sebuah pengetahuan yang dapat dimanfaatkan untuk kepentingan pemerintah dan masyarakat.



Gambar 2. Metodologi Penelitian

HASIL DAN PEMBAHASAN

Penelitian yang dilakukan melalui beberapa proses yaitu :

• PreProcessing

PreProcessing yang dilakukan yaitu proses pembersihan data mengalami tiga tahap pembersihan yaitu: *Incomplete*, *Noisy* dan *Inconsisten*. Berikut dibawah ini penjelasan dan prosesnya.

1. Incomplete

Pada tahap ini, penulis membersihkan data berdasarkan data yang tidak lengkap atau data yang tidak terisi. Kesimpulan dari tahap ini adalah pembersihan data, dalam artian

bahwa jika *attribute* data tersebut kosong atau tidak terdapat nilai didalamnya maka *attribute* data tersebut akan dihapus. Hal tersebut di atas dikarenakan penulis tidak memiliki data pendukung untuk mengisi *attribute* data yang kosong.

2. Noisy dan Inkonsistensi Data

Merupakan suatu data yang memiliki kelainan, hal ini dikarenakan karena kesalahan operator dalam memasukkan data kedalam database, permasalahan pada pentransmisian data, keterbatasan teknologi, atau tidak dilakukannya penyeragaman data, seperti data wilayah, jenis kelamin, dan lain sebagainya.

Proses pembersihan data ini menggunakan suatu tools data mining yaitu Orange.

Pengubahan Data menjadi Numeric

Pada langkah selanjutnya data yang sudah dibersihkan akan diubah value nya menjadi bentuk numeric, hal ini dimaksudkan agar pengolahan data pada MatLab dapat lebih mudah dan cepat, karena MatLab hanya memproses data-data numeric.

➤ Kategori

Pada attribute kategori data akan diubah menjadi numeric dari 1-3, berikut dibawah ini tabel pengintegrasian data pada attribute kategori.

Tabel 2 Pengintegrasian data Kategori

Data awal	Konversi data numeric
Rumah Tangga Tetap	1.00
UKM	2.00
Rumah Tangga Musiman	3.00

➤ Kabupaten

Sama seperti halnya pada langkah pertama maka data atau attribute kabupaten Cilacap akan diubah menjadi bentuk numeric. Seperti terlihat pada tabel dibawah.

Tabel 3. Transformasi data Kabupaten

Data awal	Konversi data numeric
Cilacap	1

➤ Kecamatan

Pengubahan pun berlanjut pada data atau attribute Kecamatan, hal yang dilakukan pun

sama seperti pada langkah-langkah sebelumnya.

Tabel 4. Transformasi data Kecamatan

Data awal	Konversi data numerik
Kroya	102

Disini penulis mengubah data kecamatan Kroya menjadi bentuk numeric “102” untuk memudahkan dalam pengenalan pada 17 kelurahan yang ada pada kecamatan tersebut. yang akan ditampilkan pada tabel dibawah ini.

Tabel 5. Transformasi data Kelurahan

Data awal Kecamatan	Konversi data numerik
Ayamalas	10201
Bajing	10202
Bajing Kulon	10203
Buntu	10204
Gentasari	10205
Karangmangu	10206
Karang Turi	10207
Kedawung	10208
Kroya	10209
Mergawati	10210
Mujur	10211
Mujur Lor	10212
Pekuncen	10213
Pesanggrahan	10214
Pucung Kidul	10215
Pucung Lor	10216
Sikampuh	10217

➤ Pendidikan

Pada attribute Pendidikan data akan diubah menjadi numeric dari 1-5, berikut dibawah ini tabel pengintegrasian data pada attribute Pendidikan.

Tabel 6. Transformasi data Pendidikan

Data awal	Konversi data numeric
SD	1
SLTP	2
SLTA	3
Universitas	4
Lainnya	5

➤ Pekerjaan

Pada attribute Pekerjaan data akan diubah menjadi numeric dari 1-3, berikut dibawah ini tabel pengintegrasian data pada attribute Pekerjaan.

Tabel 7. Transformasi data Pekerjaan

Data awal	Konversi data numeric
PETANI	1
SWASTA	2
Lainnya	3
PNS	4
TNI/POLRI	5

➤ Pengeluaran

Pada attribute Pengeluaran data akan diubah menjadi numeric dari 1-3, berikut dibawah ini tabel pengintegrasian data pada *attribute* Pengeluaran.

Tabel 8. Transformasi data Pengeluaran

Data awal	Konversi data numeric
500 Ribu	1
500 Ribu – 1 Juta	2
1 Juta – 1,5 Juta	3
>1,5 juta	4

➤ Masak

Pada attribute Masak data akan diubah menjadi numeric dari 1-3, berikut dibawah ini tabel pengintegrasian data pada *attribute* Pekerjaan.

Tabel 9. Transformasi data Masak

Data awal	Konversi data numeric
Minyak Tanah	1
Kayu Bakar	2
Lainnya	3

Transformasi Data

Tahapan Transformasi data adalah pengubahan format data tersimpan menjadi bentuk standar format file yang sesuai dengan aplikasi yang akan digunakan. Pada penelitian kali ini, penulis menggunakan aplikasi MatLab untuk melakukan proses data mining dengan Proses Aturan Prediksi menggunakan Regresi Linear, aturan Klasifikasi menggunakan Decision Tree. Pengubahan atau Transformasi data ke dalam format file yang sesuai adalah mengubah format file data sebelumnya yang merupakan bentuk file Tab Delimited yaitu Text atau .txt menjadi bentuk .dat.

• Proses Data Mining

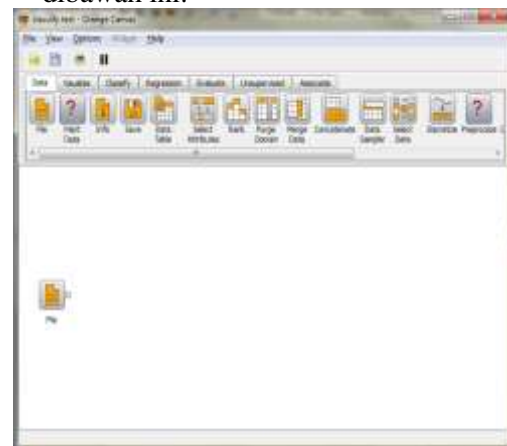
Pada tahapan ini Peneliti mencoba membuat proses aturan klasifikasi pada data mining menggunakan *decision tree*. Algoritma klasifikasi akan menggunakan data tersebut

untuk menghasilkan pengetahuan. Pengetahuan yang hendak dihasilkan dalam klasifikasi adalah untuk menggolongkan bagaimana ketentuan seorang warga mendapatkan suatu bantuan, baik itu bantuan secara materiil ataupun bantuan yang bersifat pelatihan agar warga tersebut dapat meningkatkan kesejahteraan hidupnya. Pada proses ini, Peneliti mengambil attribute kelurahan, pendidikan, pekerjaan dan pengeluaran untuk menentukan siapa warga yang memiliki ketentuan tersebut diatas berhak menerima bantuan. Berikut ini merupakan tahapan awal sampai akhir proses klasifikasi dengan *decision tree* dengan menggunakan Orange.

1. Input Data

Langkah awal proses ini adalah memasukan data yang telah dianalisa dan diproses dengan metode Cleansing (Pembersihan Data). Data tersebut merupakan data valid yang dibutuhkan untuk proses klasifikasi dengan menggunakan Orange. Berikut dibawah ini adalah langkah-langkahnya.

- Pilih Widget File pada tabulasi Data di Orange, seperti terlihat pada gambar dibawah ini.



Gambar 3. Tampilan Widget File pada Orange

Setelah Widget File berada pada kanvas, maka selanjutnya.

- klik dua kali pada widget File tersebut untuk melihat propertiesnya dan memilih file data yang terdapat pada direktori tersimpan, terlihat pada gambar dibawah ini.

The screenshot shows the 'Anzahl Kunden' report in SAP. The report is titled 'Anzahl Kunden' and is located in the 'Anzahl Kunden' section. The report displays a list of customers with the following columns: Name, Address, City, State, Country, and Date. The data is filtered by 'Anzahl Kunden' and 'Anzahl Kunden'. The report shows a total of 10 customers.

Name	Address	City	State	Country	Date
1	10	100	100	100	100
2	20	200	200	200	200
3	30	300	300	300	300
4	40	400	400	400	400
5	50	500	500	500	500
6	60	600	600	600	600
7	70	700	700	700	700
8	80	800	800	800	800
9	90	900	900	900	900
10	100	1000	1000	1000	1000

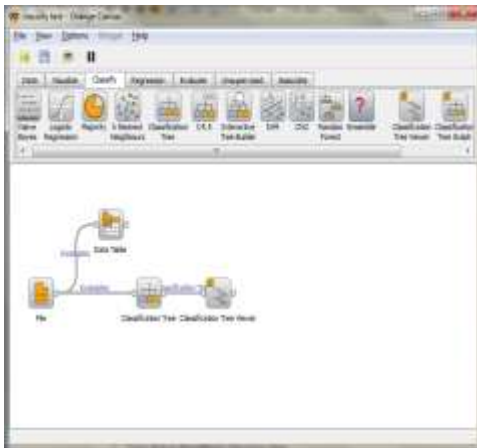
2. Lihat Data

3. Proses Klasifikasi

[illegible]

4. *View* Klasifikasi

17



Gambar 8. Tampilan Widget Classification Tree Viewer pada Orange.

Langkah selanjutnya adalah meng-klik dua kali pada widget Classification Tree Viewer tersebut untuk melihat output yang ada. output tersebut ditampilkan pada gambar dibawah ini.



Gambar 9. Tampilan Properties Widget Classification Tree Viewer pada Orange.

5. View Pohon Klasifikasi / Decision Tree

Berikut ini adalah tampilan hasil yang berbentuk suatu pohon klasifikasi dengan menggunakan widget classification tree graph yang dihubungkan dengan widget classification tree yang ada di kanvas, seperti terlihat di bawah ini.



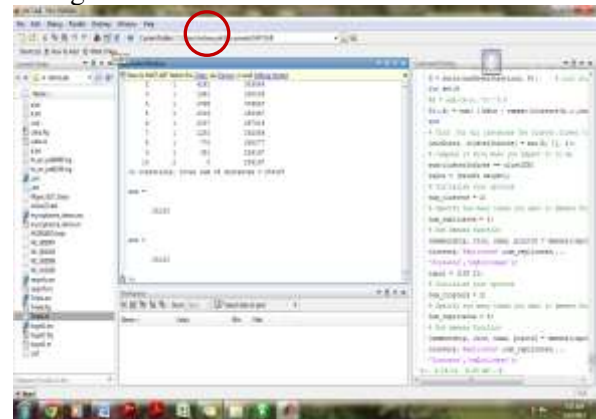
Gambar 10. Tampilan Properties Widget Classification Tree Viewer pada Orange.

Output dengan menggunakan Classification Tree graph dapat terlihat pada gambar dibawah ini.

• Proses Aplikasi

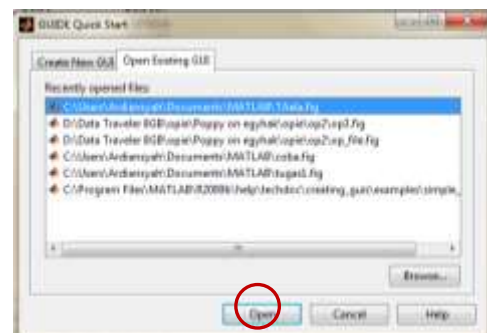
Prediksi pada Klasifikasi dengan Rumus Regresi linear untuk menentukan mana kelurahan yang memiliki prioritas untuk diberikan bantuan dari 17 kelurahan yang ada di kecamatan kroya.

1. Membuka Aplikasi berbentuk *.fig pada yang ditunjukkan lingkaran merah pada gambar di bawah ini :



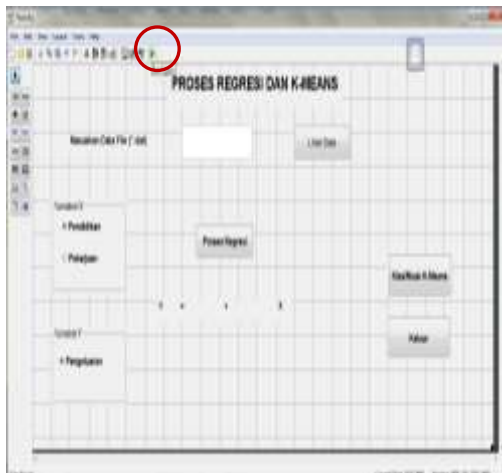
Gambar 11. Tampilan awal MatLab dengan lingkaran merah untuk membuka file Figure yang telah dibuat.

2. Kemudian akan muncul tabel dialog seperti di bawah, pilih *Open Existing GUI*. Akan tampil daftar file *.fig yang telah kita buat. Pilih file *TAela.fig*.



Gambar 12. Tampilan dialogue box open file figure

3. kemudian jalankan aplikasi dengan mengklik icon Run Figure, seperti yang ditampilkan pada gambar di bawah yang ditunjukkan oleh lingkaran merah.



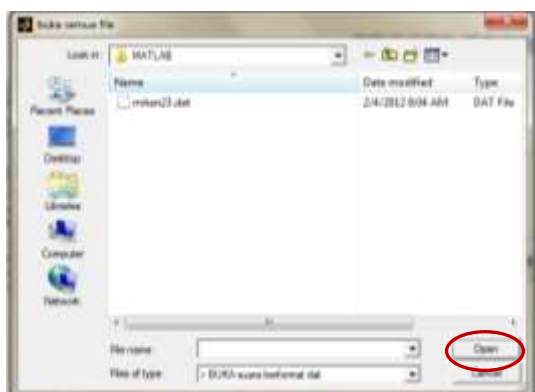
Gambar 13. Tampilan Aplikasi data Mining dengan lingkaran merah untuk menjalankan Program

4. Akan muncul *table dialog* seperti gambar di bawah, kemudian klik **Lihat Data**



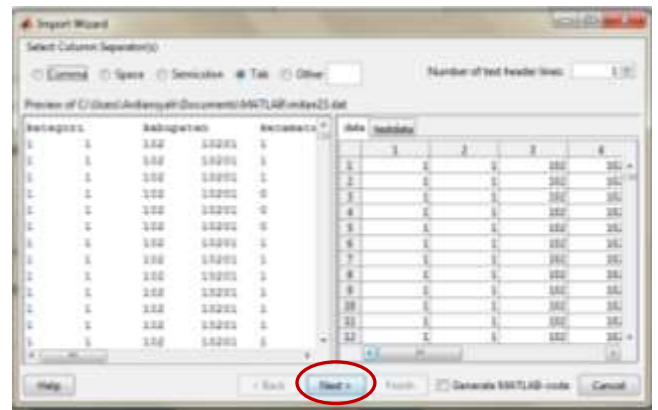
Gambar 14. Tampilan Aplikasi Data Mining untuk melihat data

5. Pilihlah file yang berbentuk *.dat yang ingin diproses. Dan klik *open*.



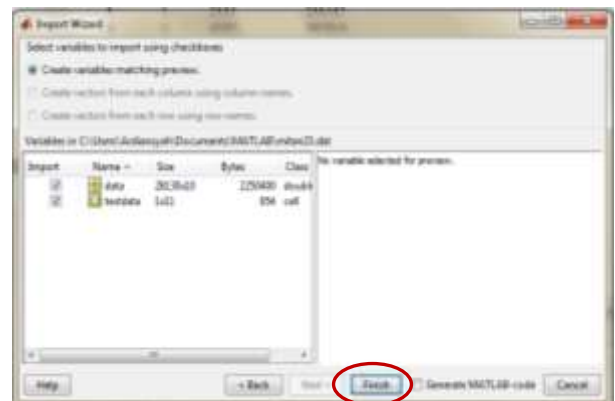
Gambar 15. Tampilan Dialogue Box Open File

6. Muncullah tabel dialog *Import Wizard* di bawah ini, klik *next*.



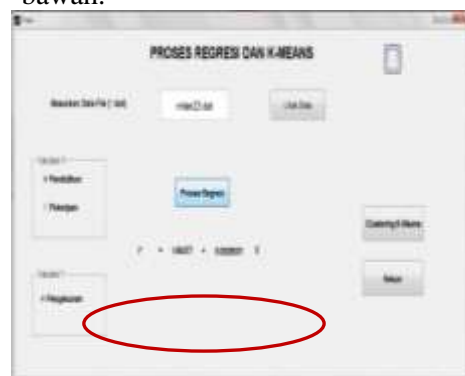
Gambar 16. Tampilan Dialogue Box Import Wizard

7. Pilih file yang ingin di import, kemudian klik *finish*.



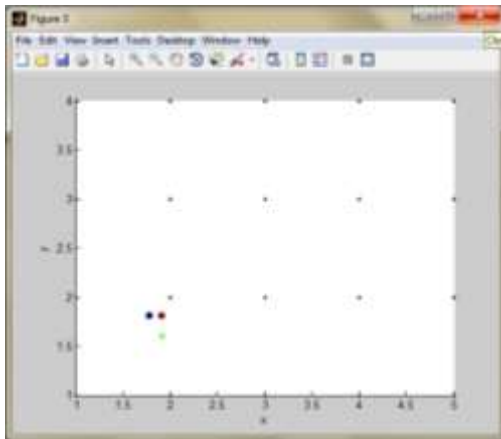
Gambar 17. Tampilan Dialogue Box Import Wizard

8. Untuk melakukan Proses Regresi terlebih dahulu pilihlah variable x dan variable y yang ingin diproses, untuk gambar di bawah variable x yang dipilih adalah pendidikan dan variable y adalah pengeluaran, sehingga akan muncul persamaan garis linear seperti yang ditunjukkan oleh lingkaran merah di bawah.



Gambar 18. Tampilan Proses Regresi

9. Proses Clustering K-means yang dilakukan yaitu menentukan secara random letak tiga buah k dan menentukan anggota yang termasuk kedalamnya. Di bawah ini adalah tampilan scatter diagram untuk k-means.



Gambar 19. Tampilan Scatter Diagram

Penarikan Informasi

Berdasarkan hasil yang didapat pada 3 langkah data mining diatas, maka dapat ditarik beberapa kesimpulan atau informasi yang dapat dijadikan pengetahuan untuk diambil suatu keputusan, informasi tersebut dapat dijabarkan sebagai berikut.

1. Klasifikasi dengan Pohon Keputusan/Decision Tree.

Berdasarkan proses software Orange dengan menggunakan widget Classification Tree Viewer dan Classification Tree Graph bahwa keputusan yang dapat diambil oleh kepala kecamatan tentang kelurahan yang akan diberikan bantuan dengan skala prioritas adalah Kelurahan Bajing Kulon, Kedawung, Pekuncen, Pesanggarahan dan seterusnya.

2. Klasifikasi dengan Regresi Linear

Untuk mendapatkan informasi yang dibutuhkan dalam pengambilan keputusan dengan menggunakan Regresi Linear maka data yang ada yaitu data Pada Kecamatan Kroya dengan 17 kelurahannya harus di pecah menjadi data per kelurahan saja. Data per Kelurahan inilah nantinya akan dihitung dengan regresi linear pada Aplikasi Data Mining yang Peneliti rancang, yang kemudian

hasil dari proses tersebut satu per satu Kelurahan yang ada di Kecamatan Kroya akan dibandingkan secara keseluruhan. Hasil dari regresi linear ini adalah mencari nilai dari Variabel Y terkecil di kelurahan yang ada.

KESIMPULAN

Dapat disimpulkan bahwa dengan menggunakan *decision tree* prioritas kelurahan yang dapat diberi bantuan yaitu: Kelurahan Bajing Kulon, Kedawung, Pekuncen, dan Pesanggarahan.

Proses pada regresi berdasarkan per kelurahan dengan atribut untuk variabel x adalah pekerjaan atau pendidikan dan atribut untuk variabel y adalah penghasilan.

Terlepas dari penjelasan pada paragraph di atas, Peneliti sadari masih banyak kekurangan dalam penelitian ini, hal ini dikarenakan karena variable data yang terlalu luas dan proses pembersihan data yang memakan waktu lama, sehingga pemilihan proses data mining dengan penggunaan algoritma yang diterapkan mungkin memiliki kelemahan dalam keakuratan hasil data yang ada, tetapi hasil yang didapatkan tidak mengurangi keakuratan algoritma atau rumus yang diterapkan.

DAFTAR PUSTAKA

- Han, Jiawei and Kamber, Micheline (2006), *Penambangan data: Concepts and Techniques*. Elsevier
- Xin Yan, & Xiao Gang Su, (2009), *Linear Regression Analysis*. London: World Scientific Publishing Co. Pte. Ltd., Covent Garden
- Larose, & Daniel T. (2005). *Discovering knowledge in data: an introduction to data mining*, John Wiley and Sons, USA
- Lior Rokach, & Oded Maimon, (2005), *Data Mining With Decision Tree*. World Scientific Publishing Co. Pte. Ltd, Series in Machine Perception Artificial Intelligence Volume 69